2002

Sample design and estimation for household surveys

Robert Graham Clark
University of Wollongong, rclark@uow.edu.au

SAMPLE DESIGN AND ESTIMATION FOR HOUSEHOLD SURVEYS

A thesis submitted in fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

from

UNIVERSITY OF WOLLONGONG

by

Robert Graham Clark

SCHOOL OF MATHEMATICS AND APPLIED STATISTICS

2002

Abstract

Household surveys are a widely-used tool for obtaining information about a population of people. A sample of households is selected followed by a sample of people within selected households. Households exhibit structure with variables measured on people in the same household often being dependent. Household sizes vary significantly and the strength of dependencies within a household may depend on its size. Traditional sample design and estimation methods ignore these dependencies.

Methodologies which explicitly allow for the dependencies which may arise within households are developed in four areas:

- estimating the design effects of standard estimators, for survey design;
- constructing new estimators to exploit dependencies within households;
- selecting a set of auxiliary variables to use in regression estimation;
- allocating the sample sizes of households and of people within households.

In each case, new methods will be developed which allow for the population structure of people within households. The new methods will be compared theo-

retically and numerically to existing methods to show whether, and under what conditions, it is worthwhile explicitly allowing for this population structure.

This thesis is concerned with the sampling error of estimators of population totals; that is, the error due to selecting only a sample and not the whole population of people. The model-assisted framework will be used.

The thesis finds that the population structure of people within households can be exploited to give several useful innovations. It is shown that, in estimating the variance at the sample design stage, the variation of household size must be considered. This variation is ignored in existing methods for estimating the design effect, and a more accurate method is developed. It is found that minor improvements can be made to standard estimators of total by considering within-household dependencies. An "integrated weighting" method, based on a linear contextual model, which has important practical advantages is found to often have slightly lower variance than non-integrated methods, contrary to common belief. Existing criteria for selecting which auxiliary variables to use in regression estimation are extended to the case of two-stage sampling, and applied to household surveys. In most household surveys, either one person or all people are selected from each selected household. More general designs, in which the number of people selected is a function of the number of people in the household, are developed. The fact that the number of people in a household is small leads to some novel and efficient sample designs and estimators.

Certification

I, Robert Graham Clark, declare that this thesis is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Robert Clark

4 March 2002

Acknowledgements

I would like to thank my friends and colleagues who helped to make my Ph.D. a fulfilling and enjoyable experience.

I am grateful to my supervisor, Professor David Steel, for his encouragement, ideas and good advice. David also gave me excellent opportunities to travel and collaborate with other researchers in survey statistics.

This research was funded by the Australian Research Council and the Australian Bureau of Statistics(ABS). Mr Frank Yu from the ABS supported my leave of absence from the ABS and provided valuable advice during the Ph.D. Both Frank Yu and Professor Alan Welsh from the University of Southampton helped to develop the research plan.

The Department of Social Statistics at the University of Southampton kindly hosted me for one year and provided an active and friendly research environment. Professor Ray Chambers made me welcome at Southampton and generously gave his time, ideas and hospitality.

Thanks to my friends at the University of Wollongong and the University of Southampton for sharing many enjoyable and challenging experiences.

Thank you Carolyn Silveri, Kerrie Gamble, Anne Owens, Kathy Hooper, Debbie Edwards and the University of Wollongong Office of Research for providing administrative support.

Thanks Bill, Stephanie, Alistair, Matt and Kathy for giving me homes away from home in Canberra. Thanks Patricia and Michelle for looking after my house.

Bill Allen, Eric Beh, Russell Familar, Michael Meagher, Nick von Sanden and Craig McLaren were a great help in proofreading the thesis.

I am grateful to my parents for their support and example.

Contents

List of Figures

List of Tables